



Recherche des successions de cultures et de leurs évolutions : analyse par HMM des données Ter-Uti en Lorraine

Marc Benoît, Florence Le Ber, Jean-François Mari

► To cite this version:

Marc Benoît, Florence Le Ber, Jean-François Mari. Recherche des successions de cultures et de leurs évolutions : analyse par HMM des données Ter-Uti en Lorraine. Agreste - les cahiers, 2001, pp.23–30. inria-00107540

HAL Id: inria-00107540

<https://hal.inria.fr/inria-00107540>

Submitted on 19 Oct 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Recherche des successions de cultures et de leurs évolutions : analyse par HMM des données *Ter Uti* en Lorraine

Marc Benoît (1), Florence Le Ber (2), Jean-François Mari (2)

(1) INRA SAD, Domaine du Joly, 88400 Mirecourt

benoit@mirecourt.inra.fr

(2) UMR 7503 LORIA, B.P. 239, 54506 Vandœuvre-lès-Nancy

leber@loria.fr, jfmari@loria.fr

Résumé

Cet article présente les résultats d'un travail de fouille de données effectué par des agronomes et des informaticiens pour extraire de la base de données *Ter Uti* lorraines des informations sur les successions de cultures pratiquées dans la région. Nous avons utilisé pour cela des modèles stochastiques développés pour l'analyse de séquences, les modèles de Markov cachés. Ces modèles permettent de représenter des observations temporelles et spatiales comme des successions d'états où les transitions entre états dépendent, suivant l'ordre du modèle, de l'état courant et des n états précédents. Ils sont donc particulièrement bien adaptés à la recherche de séquences temporelles telles que les successions culturales. Dans cet article, nous détaillons les modèles utilisés et la démarche mise en œuvre sur les données *Ter Uti* ainsi que les résultats obtenus. Nous présentons aussi les outils que nous avons développés pour faciliter la visualisation et la manipulation des résultats.

Mots-clé : Fouille de données spatio-temporelles, successions culturales, données *Ter Uti*, Lorraine, modèles de Markov cachés.

Introduction

L'occupation du territoire agricole change d'années en années, en raison de deux faits majeurs, la libération continue de territoires par la disparition d'exploitations agricoles et l'évolution des systèmes de production. Cette évolution se traduit principalement dans le changement des successions de cultures, par introduction de nouvelles cultures, raccourcissement des rotations, prise en compte de la réglementation PAC, etc.

L'étude de ces successions est un enjeu pour l'agronomie car :

- la nature de ces successions et de leur évolution est un indicateur des dynamiques en cours dans l'agriculture ;
- les successions contribuent à des effets environnementaux majeurs : pollution des ressources en eau, structuration des paysages.

Néanmoins, les successions culturales ont fait l'objet de peu d'études, et le choix à proprement parler d'une succession par un agriculteur reste un phénomène mal connu. Ce choix est révélateur du métier d'agriculteur et intègre de nombreux facteurs : dates de récolte et d'implantation des cultures, état laissé par une culture après la récolte et organisation de chantiers entre parcelles, etc.

Les données *Ter Uti* sont un moyen d'accéder – au moins à un niveau global – à cette connaissance. Le but de l'étude que nous avons menée est en effet de recenser les successions dominantes et leurs évolutions à l'échelle de toute une région et des petites régions agricoles qui la composent. Cette information pourra être ensuite confrontée aux données issues d'enquêtes de terrain, auprès des agriculteurs et techniciens agricoles et pourra également servir de base aux études sur l'évolution de l'occupation des sols et sur ses effets environnementaux.

Cette étude a été menée conjointement par des informaticiens “fouilleurs de données” et des agronomes, et s’est déroulée en plusieurs temps :

1. étude des répartitions de cultures au cours de la période 1992-99,
2. recherche des successions,
3. étude de l’évolution des successions et de leur répartition spatiale.

L’article s’organise de la façon suivante : la première partie est un exposé de la problématique agronomique, la deuxième partie présente le principe de la méthode et la base de données dont nous disposons, la troisième partie décrit les différents modèles et traitements que nous avons utilisés, la quatrième partie présente des résultats. La dernière partie de l’article est une discussion.

1 Les successions culturelles, concept central en agronomie

L’agronomie organise ses progressions méthodologiques autour du concept de système de culture [7, 15]. Le système de culture est ainsi défini comme une portion de territoire conduite de façon homogène, où on peut définir les cultures qui s’y succèdent et les itinéraires techniques qui leur sont appliqués. Les successions culturelles sont donc l’un des deux pôles organisateurs des systèmes de culture.

Pour les agriculteurs, le choix des successions culturelles est le révélateur a posteriori des ajustements qu’ils réalisent entre leurs parcelles disponibles et l’importance surfacique des couverts végétaux qu’ils ont choisis. Une lecture de ces successions nous renseigne sur les difficultés et les simplifications auxquelles sont confrontés les agriculteurs dans leur ajustement “territoire – couverts végétaux”.

D’autre part, les successions culturelles sont des résultats d’organisation d’agriculteurs qui influencent fortement les effets de l’agriculture sur les ressources renouvelables :

- l’érosion (proportion des inter-cultures longues),
- la dégradation des qualités d’eaux souterraines,
- les monotonies paysagères.

Pour l’agronome, disposer des successions culturelles à une échelle régionale permet de mieux choisir ses dispositifs. Qu’il s’agisse d’enquêtes ou de mise en place d’expérimentations, cet outil permet d’instruire le domaine de validité et d’extrapolation de ses études.

2 La méthode utilisée : fouille de données par modèles de Markov

À l’heure actuelle, on définit la *fouille ou extraction de connaissances à partir de données* comme une activité consistant à analyser un ensemble de données brutes de façon à en extraire des informations qui peuvent être considérées comme des éléments de connaissances et donc devenir exploitables [9]. Un système de fouille de données s’articule généralement autour de quatre composantes principales : (1) les bases de données et leur système de gestion, (2) un système à base de connaissances d’aide à la résolution de problème sur le domaine relatif aux données, (3) un système d’étude et d’analyse de données symboliques, s’appuyant sur l’induction ou la classification, un système d’analyse de données numériques et de statistiques, (4) une interface se chargeant des interactions et de la visualisation des résultats intermédiaires et finaux.

Nos travaux ont porté pour le moment sur les éléments (1), (3) et (4) du système idéal décrit ci-dessus. La base de données contient les données *Ter Uti* pour la Lorraine de 1992 à 1999. Le système d’analyse que nous proposons est fondé sur des méthodes de classification non supervisée et d’apprentissage à l’aide de modèles de Markov cachés (HMM pour Hidden Markov Model). Ce sont en effet des outils appropriés pour dégager des régularités temporelles ou spatiales comme le montrent les travaux en reconnaissance de la parole [8, 13] et dans différents domaines : segmentation d’images, bio-informatique, robotique, etc. Nous avons développé également des outils de visualisation pour permettre la manipulation des données et des résultats du système d’analyse.

2.1 La base de données

Notre base de données est constituée de l'enquête *Ter Uti* Lorraine : nous disposons de résultats sur 23756 points pour les années 1992 à 1999. Les points sont situés régulièrement sur des photos aériennes, elles-mêmes réparties dans le territoire, de telle sorte que la représentativité d'un point est proche de 100 hectares (voir [11] pour les détails sur la "grille" *Ter Uti*). Outre la séquence temporelle des occupations de chaque point, nous connaissons ses voisins, c'est-à-dire que nous connaissons la disposition relative de chaque point et de chaque photo aérienne. Les occupations sont réparties en différentes classes (environ 80) dont certaines ne sont pas ou peu présentes en Lorraine ("glaciers", "neiges éternelles" mais aussi "pomme de terre") aussi avons nous restreint le nombre de classes à 49, par regroupement ou suppression. Parmi ces classes, nous nous intéressons particulièrement aux occupations agricoles dominantes en Lorraine, c'est-à-dire : maïs, blé, orge, colza, prairies temporaires, prairies permanentes, vergers.

Le but de l'étude est de trouver les régularités spatiales et temporelles qui rendent compte de l'organisation – successions et assolements – de l'agriculture lorraine. Nous admettons pour cela l'hypothèse que, si les règles de successions adoptées par les agriculteurs ne sont pas connues avec précision, en revanche, sur une parcelle, la règle de succession ne dépend que de l'occupation actuelle de la parcelle et de l'occupation de la ou des deux années précédentes. Cette hypothèse nous permet d'utiliser les modèles de Markov d'ordre un et deux pour traiter les données dont nous disposons.

2.2 Définition d'un HMM

Les modèles de Markov cachés dérivent des chaînes de Markov. La chaîne de Markov définit un seul processus stochastique : elle possède un ensemble d'états – par exemple les cultures successives d'une parcelle – directement observables. Dans un HMM, une observation n'est pas uniquement associée à un état mais dépend d'une densité de probabilité définie sur chacun des états. Le HMM permet de représenter deux processus stochastiques, le premier gouvernant le second [1] :

- le premier processus est défini sur un ensemble d'états et caché pour un observateur. C'est une chaîne de Markov d'ordre un ou deux. On considère ainsi que la répartition des cultures dans une région donnée évolue selon un processus de Markov. La répartition à un pas de temps donné ne dépend que de la répartition aux pas précédents suivant l'ordre du modèle.
- le deuxième processus est qualifié de visible. Il émet une observation à chaque pas de temps en fonction des densités de probabilité définies sur chacun des états. On suppose ainsi que l'observation d'une culture une année donnée sur une parcelle suit la loi définie sur l'état atteint cette année par le processus caché.

Finalement, un modèle de Markov caché est défini par la donnée de :

- $S = \{O, s_1, s_2, \dots, s_N, F\}$, un ensemble fini comprenant $N + 2$ états, dont un état initial O et un état final F ;
- \mathcal{A} la matrice donnant les probabilités de transition entre états :
- $\mathcal{A} = (a_{ij})$ pour un modèle d'ordre 1 (HMM1),
- $\mathcal{A} = (a_{ijk})$ pour un modèle d'ordre 2 (HMM2) ;
- $b_i(\cdot)$ les lois des densités associées aux états s_i .

La matrice \mathcal{A} est initialisée avec un ensemble de valeurs qui permettent de définir la topologie du graphe des transitions entre états : quelles sont les transitions autorisées, aller simple ($a_{ij} = 1$, $a_{ji} = 0$) aller-retour ($a_{ij} = a_{ji} = 1$), bouclage (a_{ii} non nul), etc. (voir figure 1 pour un exemple de graphe).

2.3 Estimation automatique d'un HMM

Une fois donné un corpus de données et la topologie du graphe des transitions entre états, différents algorithmes permettent l'apprentissage d'un HMM. Quelque soit l'ordre des modèles, nous utilisons l'algorithme Forward - Backward [12] qui est une variante de l'algorithme EM [6]. L'apprentissage se fait itérativement en partant d'un modèle où toutes les transitions sont équiprobables et où les densités $b_i(\cdot)$ sont fixées. L'algorithme Forward - Backward calcule un nouveau modèle plus adapté aux données dans lequel la vraisemblance du corpus a augmenté. Ce nouveau modèle est utilisé dans une nouvelle itération jusqu'à ce que la vraisemblance du corpus atteigne un maximum local. Le résultat est constitué par les nouvelles valeurs des transitions a_{ijk} et des densités $b_i(\cdot)$ (cf. figure 2). Lorsque N est le nombre d'états et T le nombre d'observations, l'algorithme Forward - Backward a une complexité en N^3T pour un HMM2.

Le choix du modèle initial influe sur le résultat final. Pour évaluer l'adéquation du modèle obtenu par convergence, nous utilisons une mesure de distance entre états [16]. Par essais successifs, nous définissons et obtenons un modèle dépendant de plusieurs facteurs : résultats des expériences précédentes, topologie initiale, mode de convergence, critère d'arrêt.

3 Modèles et traitements

3.1 Étude des répartitions de cultures dans le temps

Nous nous sommes tout d'abord intéressés à la définition de périodes d'observation pendant lesquelles la distribution des cultures ne varie pas. Pour la période étudiée (1992-99) nous limitons donc à des modèles possédant deux ou trois états mais autorisant des transitions en boucle sur les états comme le montre la figure 1. Ce modèle permet de segmenter la période étudiée en trois sous-périodes d'environ trois ans où le système est supposé stationnaire. On justifie a posteriori le découpage en deux ou trois états sur 8 ans par le fait que trois ans correspond à l'échelle de temps d'une rotation en Lorraine.

La figure 2 illustre un résultat de classification dans laquelle nous avons cherché trois situations stables de production agricole. Entre les années 92 et 99, la région étudiée (PRA¹ 306) est passée par trois états de distributions de cultures différentes. Dans ces distributions, on voit immédiatement l'importance des bois et prairies (de l'ordre de 60 %), la progression du blé et du colza au détriment de l'orge, et la disparition de la jachère au fil des ans. À partir de ces résultats, nous avons déterminé les occupations à examiner prioritairement dans l'ensemble des occupations de la base *Ter Uti* ; c'est-à-dire à la fois les occupations les plus fréquentes et les plus instables *a priori* : blé, orge, maïs, colza et prairies.

3.2 Étude de successions de cultures

Dans un HMM tel que défini dans le modèle 1, il n'est pas possible de mesurer la probabilité d'une succession de cultures puisqu'une culture n'apparaît qu'à l'intérieur d'une répartition constituant l'état. Pour pallier ce défaut, nous introduisons des états uniquement associés aux cultures majoritaires que nous voulons étudier (blé, maïs, orge, ...). Le HMM obtenu possède deux types d'états : les états "de réserve" qui sont associés normalement à des répartitions de cultures – comme dans le modèle 1 – et les états "de Dirac" qui sont associés à une culture seule et définis par une densité où la probabilité de cette culture vaut 1 et les probabilités de toutes les autres occupations valent 0. Les états de réserve permettent de faire une segmentation progressive des données. Les cultures qui apparaissent comme étant les plus fréquentes dans un état de réserve, ou qui nous intéressent particulièrement, peuvent être placées dans un nouvel état de Dirac dans une expérience suivante. Finalement la figure 3 donne la topologie d'un modèle 2 qui sera utilisé

1. PRA pour Petite Région Agricole.

par la suite. Ce modèle permet d'étudier les transitions entre les états associés au blé, orge, maïs, colza et prairie. Les états 2, 3 et 4 jouent le rôle d'états de réserve.

Les résultats obtenus par le modèle 2 sont visualisés au moyen d'un graphe tel que présenté en figure 4 pour la PRA 306. Les transitions entre états de Dirac sont représentées par des lignes brisées dès que leur probabilité atteint un certain seuil. L'agronome en fait l'interprétation suivante.

Nous voyons sur ce graphique deux ensembles de couverts pérennes qui n'échangent aucune surface avec les autres couverts sur toute la période : les prairies permanentes, les bois ; de plus apparaissent deux petits groupes de monocultures : la monoculture de blé qui reste stable sur la période, et la monoculture de maïs qui diminue en fin de période.

L'orge semble changer de statut dans les successions culturales : d'un rôle de "tampon" à celui d'élément stable d'une rotation :

- au début de la période, l'orge précède un colza ou un maïs ou même un blé et existe en couple orge-orge,
- à la fin de la période, se lit une forte régularité blé-orge-colza, où l'orge n'est plus suivie que de colza.

L'orge permet en effet une préparation précoce du sol pour le colza car sa récolte est tôt en saison.

A contrario, le blé reste à une place semblable dans la succession : il est précédent à maïs, colza, orge ... ou à lui-même.

À partir de ces résultats, nous nous sommes intéressés à l'évolution et aux transitions entre successions avec pour hypothèse, issue de l'interprétation de la figure 4, une simplification des types de successions. Par ailleurs nous présentons au paragraphe 4 d'autres résultats du modèle 2, mis en œuvre sur différents jeux de données (représentant différentes petites régions).

3.3 Étude des évolutions de successions

Nous avons alors tenté de déterminer les successions majoritaires en fixant leur taille à 3. Cette valeur n'est pas arbitraire mais correspond à une observation sur la Lorraine : dans la presque totalité des cas, les successions s'organisent en rotation avec des têtes de rotation qui reviennent au plus tous les trois ans. Tous les triplets possibles de la base sont considérés (6 triplets pour 8 années, donc 81915 pour 23756 points, mais seulement 1109 triplets différents). Chaque triplet constitue une observation de trois occupations successives faites pendant la période 1992-99. Les résultats de la classification ont été étudiés sous forme de simples tableaux et ont permis de :

- repérer les successions dans les triplets, c'est-à-dire regrouper les différentes permutations : on vérifie ainsi que les triplets (colza blé orge), (blé orge colza) et (orge colza blé) ont à peu près la même représentation dans chacun des états ; de même pour (blé colza blé) et (colza blé colza) ;
- repérer et évaluer les successions majoritaires : on vérifie que deux successions (colza blé orge) et (colza blé) représentent une grosse partie des terres cultivées (environ 28 %) ;
- étudier la progression, l'apparition ou la disparition des différentes successions dans la période considérée.

À l'issue de cette analyse, nous avons défini les successions à étudier davantage sous forme de grandes classes ou de rotations types (colza+2céréales, colza+1céréale, maïs+2céréales, maïs+1céréale, monocultures). Nous avons alors, comme lors de la première étape, construit un nouveau modèle – appelé modèle 3 – dans lequel des états n'émettent que ces triplets et leurs permutations circulaires (cf. figure 5).

Les résultats de ce modèle peuvent être présentés de la même façon que sur la figure 4 : les trajectoires entre états mises en évidence dans la figure 4 deviennent des lignes droites dans les figures 6 et 5. Ces deux figures représentent les résultats obtenus sur la même région (PRA 306), avec deux seuils différents : sur la figure 6 ce seuil est fixé à 1 % et à 0,4 % sur la figure 5. En effet, l'affichage des résultats peut être modifié par l'utilisateur, et donc être plus ou moins lisible, et faire apparaître plus ou moins d'informations.

Ainsi, l'agronome va interpréter différemment les deux figures 6 et 5. Il a beaucoup de mal de travailler au seuil de 0,4 %. En revanche, avec le seuillage à 1,0 %, il trouve des tendances très

instructives :

- la classe des monocultures de prairies domine largement toutes les autres et reste stable (aucune évolution de type retournement de prairies ou remise en herbe de culture n’apparaît) ;
- l’état de réserve diminue pour alimenter les successions colza+2céréales et maïs+2céréales ;
- les transitions sont nombreuses entre colza+céréale et colza+2céréales, ce qui signifie que sur une même parcelle ces deux successions peuvent avoir lieu (de plus en plus vrai plus on arrive en fin de période) ;
- les monocultures de céréales sont peu nombreuses, instables dans le temps, et non reliées à d’autres successions ;
- la succession maïs+2céréales est très stable et augmente légèrement en fréquence au cours de la période.

En conclusion, cette région reste à dominance de monoculture prairiale, avec une tendance à accroître les successions (colza ou maïs)+2céréales dans les parcelles cultivées.

4 Exemples

Nous donnons quelques exemples des successions trouvées sur deux petites régions agricoles (outre la PRA 306 présentée ci-dessus) et la Lorraine entière. Les graphiques issus du traitement par le modèle 2 sont analysés.

4.1 PRA 308 (figure 8)

Sur ce graphique apparaissent nettement deux simplifications :

- une diminution du nombre des successions : leur nombre total diminue, et les plus fréquentes (colza-blé-orge) deviennent plus dominantes,
- une diminution des monocultures : les monocultures de maïs disparaissent et ne subsistent que les monocultures de céréales, orge ou blé.

Ce graphique donne l’impression de plus de “facilité” en fin de période : diminution des monocultures, et dominance accrue de la succession colza-blé-orge, avec forte augmentation d’une succession “raccourcie” : colza-blé-colza-blé.

4.2 PRA 309 et 310 (figure 9)

Ici, l’organisation peu structurée des successions culturelles donne l’image d’une région où la recherche de solutions stables est difficile :

- les monocultures dominent nettement : prairies permanentes et maïs, ceci dans une matrice très boisée.
- les successions de culture annuelles sont erratiques : la succession blé-maïs-orge apparaît comme un pôle de stabilité, mais la succession maïs-orge-orge-maïs existe aussi. Cette instabilité est renforcée par les transitions entre prairies permanentes et l’état de réserve (qui représente tous les autres couverts), et entre prairies permanentes labourées pour devenir des maïs.
- le regroupement des céréales (blé et orge) dans une même catégorie simplifierait peut-être le graphe. En tout état de cause, il s’agit d’une région très différentes des précédentes.

4.3 Lorraine entière (figure 10)

Pour l’ensemble de la région Lorraine, le commentaire est le même que pour la PRA 306 (figure 4). Toutefois, deux nuances sont à souligner : les prairies sont moins abondantes et les forêts plus présentes en Lorraine que dans la région 306, et au cours de la période, les monocultures d’orge s’interrompent deux ans plus tôt dans la région 306 qu’en Lorraine. Nous pouvons cependant poser que sur la période 1992 - 1999, la région 306 est très représentative des tendances d’évolution des successions culturelles existant en Lorraine.

Nous pouvons de la même manière évaluer la ressemblance entre les successions de Lorraine et celles des différentes PRA, telles qu’elles apparaissent à travers les représentations graphiques que nous construisons à partir du traitement des données *Ter Uti* par les HMM. Finalement ces représentations permettent d’identifier la représentativité d’une PRA vis-à-vis de l’ensemble de la région Lorraine.

5 Segmentation spatiale

Jusqu'à présent, nous avons classé les points en fonction de leurs prédécesseurs et successeurs temporels régis par une chaîne de Markov d'ordre deux. Le processus de classification n'utilise pas l'information spatiale de la base de données *Ter Uti*. Dans un espace de dimension deux, comme une image, la notion de chaîne de Markov se généralise en donnant naissance à la notion de champ de Markov : l'état d'un point (appelé un site) dépend seulement de ses voisins spatiaux. Les algorithmes d'estimation et de classement deviennent alors plus complexes à mettre en œuvre et à ce jour, aucune solution informatique a montré son efficacité. Au contraire, en suivant l'exemple de [2], nous avons introduit une relation d'ordre sur les points du plan qui respecte autant que possible la notion de voisinage : on passe alors d'un problème de segmentation sur deux dimensions à un problème de segmentation sur une seule dimension, comme précédemment. Les HMM permettent ainsi d'effectuer une segmentation en régions géographiques homogènes du point de vue des densités spatiales de cultures. Cette méthode a déjà montré des résultats comparables aux méthodes fondées sur les champs de Markov dans le cas de segmentation d'images [2].

Pour ordonner les points *Ter Uti*, nous utilisons une courbe de Peano qui est une courbe fractale et parcourt tous les points du plan en respectant la notion de voisinage spatial. Nous l'avons adaptée pour parcourir l'ensemble des données en respectant les deux niveaux de résolution spatiale propres aux données *Ter Uti* : un premier parcours permet d'introduire une relation d'ordre sur les 32x32 photos aériennes couvrant la Lorraine² ; un deuxième parcours à l'intérieur des photos aériennes ordonne les sites *Ter Uti*. Nous pouvons ainsi réaliser une segmentation spatiale des données : un point n'est plus classé en fonction de ses précédents temporels, mais en fonction de ses précédents spatiaux selon la courbe de Peano.

Avec un changement minime dans le programme d'apprentissage, nous avons traité l'ensemble des données, chaque année individuellement et toutes les années ensemble. Nous avons traité les données brutes (un point = une culture) et les données des successions (un point = une succession de cultures sur trois années). Ce sont ces dernières qui ont fourni les résultats les plus stables. Ce processus de classification est non supervisé ; l'utilisateur définit seulement le nombre initial d'états correspondant au nombre maximum de régions à découvrir. Nous obtenons des cartes (figure 11) où se distinguent cinq états principaux, les autres états provenant d'une division abusive par l'algorithme :

- un état – état 8 – à majorité de bâti (30 %), forêt, sols nus, zones humides, qui suit le cours des grandes vallées ;
- un état – état 9 – à majorité de forêt (98 %) qui englobe le massif vosgien et les forêts de la Meuse ;
- un état – état 2 – à majorité de prairies (30 %), forêt (20 %) et cultures fourragères (6 %) caractéristique des régions d'élevage ;
- un état – état 3 – à majorité de cultures (30 % pour blé, colza, et orge) et prairie (10 %) définissant les régions à dominante céréalière ;
- un état – état 7 – à majorité de prairies (68 %) et prés vergers (5 %) au fond des vallées vosgiennes et en pieds de côtes.

On retrouve globalement sur la figure 11 la localisation des occupations affichées sur la carte de la figure 12 telles qu'elles sont établies à partir de données satellitaires à une échelle 4 fois plus fine (programme *Corin Land Cover*). La localisation des états est interprétable en fonction de la géologie et des caractéristiques techniques des exploitations. Ainsi on retrouve les régions géologiques des Vosges gréseuses, des Vosges granitiques, du plateau lorrain (calcaire), des vallées de la Moselle et de la Meuse, des vallées vosgiennes, des plaines argileuses. On trouve aussi la distinction, sur le même terrain géologique (argiles de Keuper), entre les petites exploitations laitières du Châtenois et les grandes structures mixtes du Saulnois. Cette dernière distinction, méritera d'être confortée par une étude sur les types d'exploitations agricoles en fonction des petites régions agricoles.

2. Nous avons ajouté des photos "hoires" pour obtenir un espace carré, nécessaire au parcours de Peano1.

Discussion et conclusion

Les allers et retours entre agronomes et informaticiens se sont traduits par l'élaboration de plusieurs modèles pour valider les hypothèses des agronomes. Nous nous sommes aidés de la durée des états comme révélatrice de l'instabilité des successions ainsi que du contenu de l'état de réserve qui capte toutes les exceptions et permet d'effectuer une segmentation progressive des données. À partir d'un modèle simple – le modèle 1 – donné dans la figure 1, nous avons élaboré les modèles 2 et 3 pour mesurer plus précisément des choix de successions. La figure 5 montre un système quasiment en équilibre. Les successions dominantes ont été trouvées, elle sont représentées par les états faiblement connectés entre eux. Ces traitements peuvent être effectués à différentes échelles, PRA ou région entière, pour peu que le nombre de points soit suffisants. Les résultats conduisent à des analyses et des comparaisons très intéressantes pour les agronomes, comme montré au paragraphe 4.

La carte issue d'une segmentation 2D est un document porteur d'une information de synthèse appréciée des agronomes qui la comparent spontanément à d'autres cartes et ainsi la valident en partie. Ces résultats permettent de remettre à jour, confirmer ou infirmer les connaissances sur la région étudiée. En découle également la spécification d'une nouvelle classification dans laquelle nous devons croiser différentes sources de données, les données *Ter Uti*, les données issues d'enquêtes auprès des agriculteurs et relatives au type d'exploitation agricole mais aussi les cartes de différentes données pedo-morphologiques.

Enfin, il faut souligner que les successions culturelles sont des objets de recherche centraux [14] révélateurs du métier d'agriculteur, mais qui ont donné lieu à peu de publications. L'intérêt de notre étude est donc double : d'une part apporter des informations générales qui valident et complètent les informations connues sur le terrain ; d'autre part construire des outils pour avoir "en routine" des informations sur les successions culturelles, à partir de bases de données statistiques telles que *Ter Uti*. Ainsi, les algorithmes développés sur les données lorraines ont pu être utilisés sur les données du bassin de la Seine [5, 4]. Finalement ces connaissances pourront entrer dans les modèles actuellement en cours de développement [10, 3] qui permettront d'effectuer des simulations prospectives sur l'occupation agricole du sol et ses effets en matière d'environnement.

Remerciements

Nous remercions le service régional des statistiques agricoles de la DRAF Lorraine pour l'accès aux données *Ter Uti*.

Références

- [1] Baker (J. K.). – Stochastic Modeling for Automatic Speech Understanding. *In : Speech Recognition*, éd. par Reddy (D.R.), pp. 521 – 542. – New-York, Academic Press, New York, 1974.
- [2] Benmiloud (B.) et Pieczynski (W.). – Estimation des paramètres dans les chaînes de Markov cachés et segmentation d'images. *Traitement du signal*, vol. 12, n5, 1995, pp. 433 – 454.
- [3] Benoît (M.) et Papy (F.). – Pratiques agricoles et qualité de l'eau sur le territoire alimentant un captage. *In : L'eau dans l'espace rural*, pp. 323–338. – INRA, 1997.
- [4] Bornerand (C.). – Dynamique des pratiques culturelles dans le bassin de la marne depuis les années 70. – Mémoire de DAA ENSAIA, septembre 2000. INRA SAD Mirecourt.
- [5] Caty (M.). – Évolution des pratiques agricoles et liens avec l'évolution de la qualité de l'eau dans le bassin de la Seine. – Mémoire de fin d'étude ENGEES, 1999. INRA SAD Mirecourt.

- [6] Dempster (A.P.), Laird (N.M.) et Rubin (D.B.). – Maximum-Likelihood From Incomplete Data Via The EM Algorithm. *Journal of Royal Statistic Society, Ser. B (methodological)*, vol. 39, 1977, pp. 1 – 38.
- [7] Gras (R.), Benoît (M.), Deffontaines (J.-P.), Duru (M.), Lafarge (M.), Langlet (A.) et Osty (P.-L.). – *Le fait technique en agronomie. Activité agricole, concepts et méthodes d'étude.* – INRA - L'Harmattan, 1989. 160 pages.
- [8] Jelinek (F.). – Continuous Speech Recognition by Statistical Methods. *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 64, n4, April 1976, pp. 532 – 556.
- [9] Kodratoff (Y.). – L'extraction de connaissances à partir de données : un nouveau sujet pour la recherche scientifique. In : *Actes des journées Ingénierie des Connaissances et Apprentissage Automatique*, éd. par Charlet (J.). pp. 539–566. – Roscoff, 1997.
- [10] Le Ber (F.) et Benoît (M.). – Modelling the spatial organisation of land use in a farming territory. Example of a village in the Plateau Lorrain. *Agronomie: Agriculture and Environment*, vol. 18, 1998, pp. 101–113.
- [11] Ledoux (M.) et Thomas (S.). – De la photographie aérienne à la production de blé. *AGRESTE, la statistique agricole*, no5, 1992.
- [12] Mari (J.-F.), Haton (J.-P.) et Kriouile (A.). – Automatic Word Recognition Based on Second-Order Hidden Markov Models. *IEEE Transactions on Speech and Audio Processing*, vol. 5, janvier 1997, pp. 22 – 25.
- [13] Mari (J.-F.) et Napoli (A.). – Modèles stochastiques pour la classification de signaux temporels. In : *Actes des cinquièmes rencontres de la société francophone de classification*, pp. 51 – 54. – Lyon, France, Septembre 1997.
- [14] Sebillotte (M.). – Les systèmes de culture. In : *Encyclopedia Universalis.* – 1988.
- [15] Sebillotte (M.). – Systems Research and Action. Interdisciplinary Excursions. In : *Systems-Oriented Research in Agriculture and Rural Development. Lectures and debates (Int. Symposium - Montpellier, France, 21-25 Nov. 1994).* pp. 35–72. – CIRAD, 1996.
- [16] Tou (J. T.) et Gonzales (R.). – *Pattern Recognition Principles.* – Addison-Wesley, 1974.

Figures

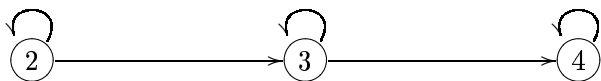


FIG. 1 – *Modèle 1 : modèle gauche - droite à trois états. Les transitions en boucle sont autorisées, le système peut rester dans le même état pendant une période mais ne peut pas revenir en arrière.*

état 2		état 3		état 4	
bois	0,296	bois	0,297	bois	0,295
prairies	0,296	prairies	0,264	prairies	0,263
blé	0,084	blé	0,115	blé	0,105
orge	0,051	colza	0,065	colza	0,077
maïs	0,047	maïs	0,047	orge	0,050
colza	0,047	orge	0,043	maïs	0,047
jachères	0,020	jachères	0,009	jachères	0,007

FIG. 2 – *Modèle 1 : Segmentation des données de la PRA Plateau lorrain sud en trois périodes pendant lesquelles les observations sont supposées stationnaires. Les états sont dénotés 2, 3 et 4.*

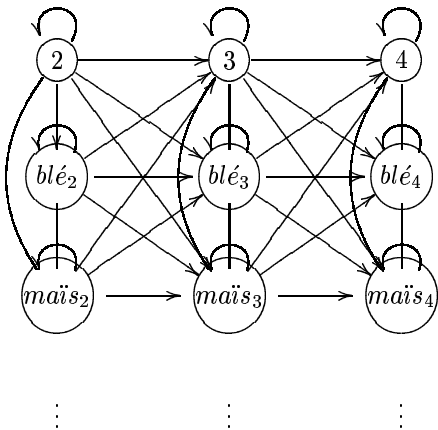


FIG. 3 – *Modèle 2 : les états notés 2, 3 et 4 sont associés à une distribution de cultures, contrairement aux états dénommés par une occupation. Le nombre de colonnes définit le nombre de périodes d’observation. Les connexions sans flèches représentent des transitions bidirectionnelles.*

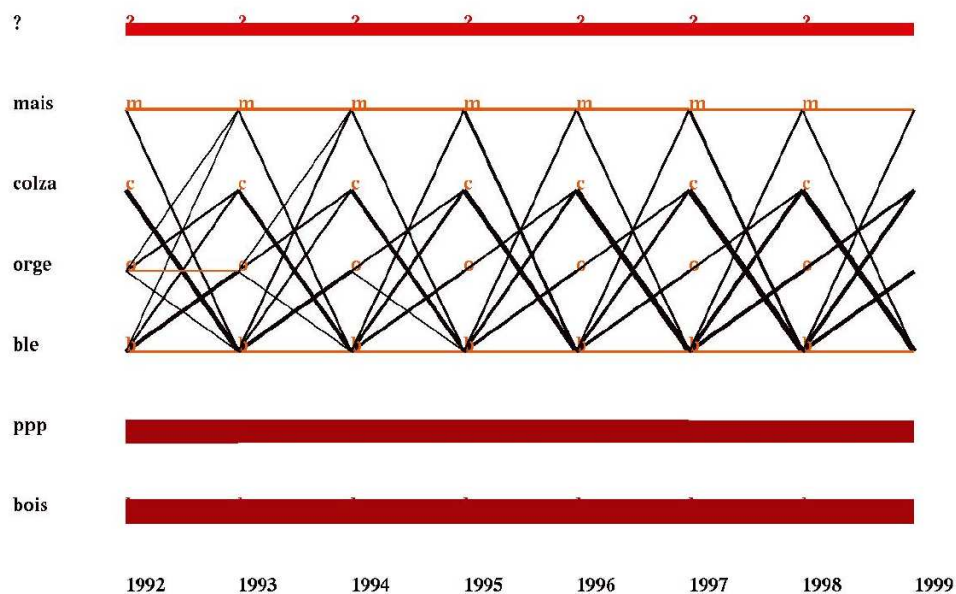


FIG. 4 – Résultats du modèle 2 sur les données de la PRA Plateau lorrain sud (PRA 306) entre 1992 et 1999 : les probabilités de transitions d'un état à l'autre sont représentées en fonction du temps dès qu'elles dépassent 1,0 %. L'épaisseur du trait est proportionnelle à la valeur de la probabilité. La ligne dénotée “?” correspond à un état de réserve affiché.

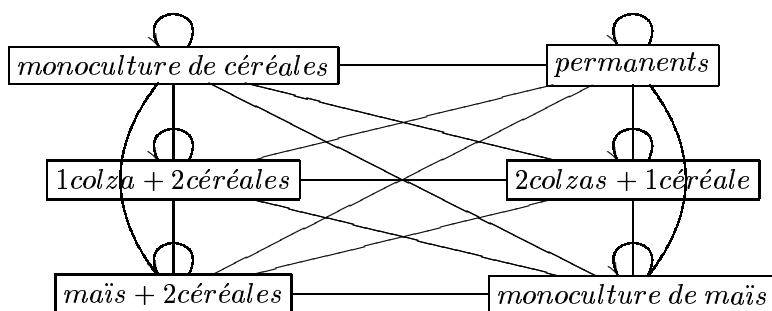


FIG. 5 – Modèle 3 : modèle de triplets, similaire au modèle 2. Un état particulierise une rotation type. Toutes les transitions sont bidirectionnelles.

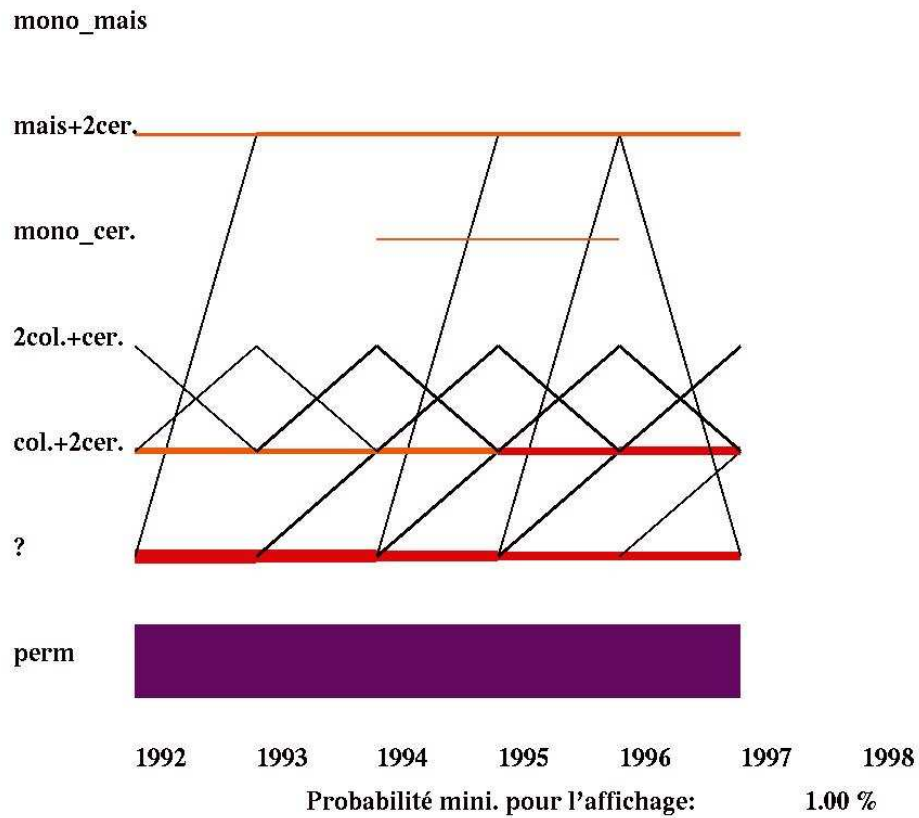


FIG. 6 – Résultats obtenus à partir du modèle 3 sur les données de la PRA Plateau lorrain sud entre 1992 et 1997. Une observation est une rotation type. Les transitions entre rotations sont indiquées par les lignes obliques, les variations de l'importance d'une même rotation sont représentées par l'épaisseur du trait horizontal, dès que la probabilité dépasse 0,4 %.

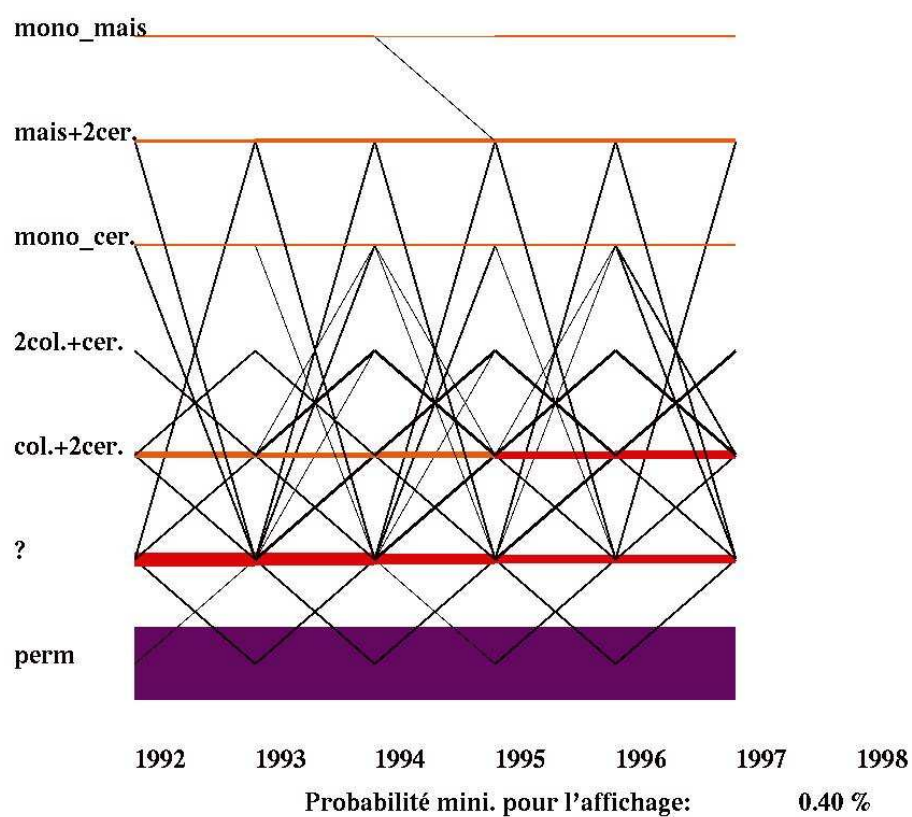


FIG. 7 – Résultats du modèle 3 sur les données de la PRA Plateau lorrain sud entre 1992 et 1997 : les probabilités de transition sont représentées dès qu'elles dépassent 1,0 %.

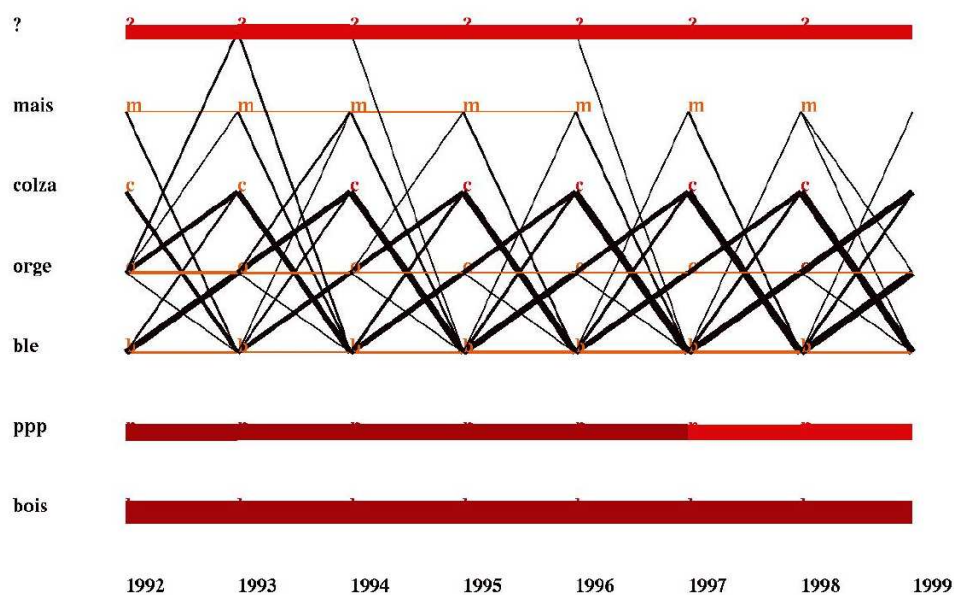


FIG. 8 – Résultats du modèle 2 sur les données de la PRA 308 entre 1992 et 1999.

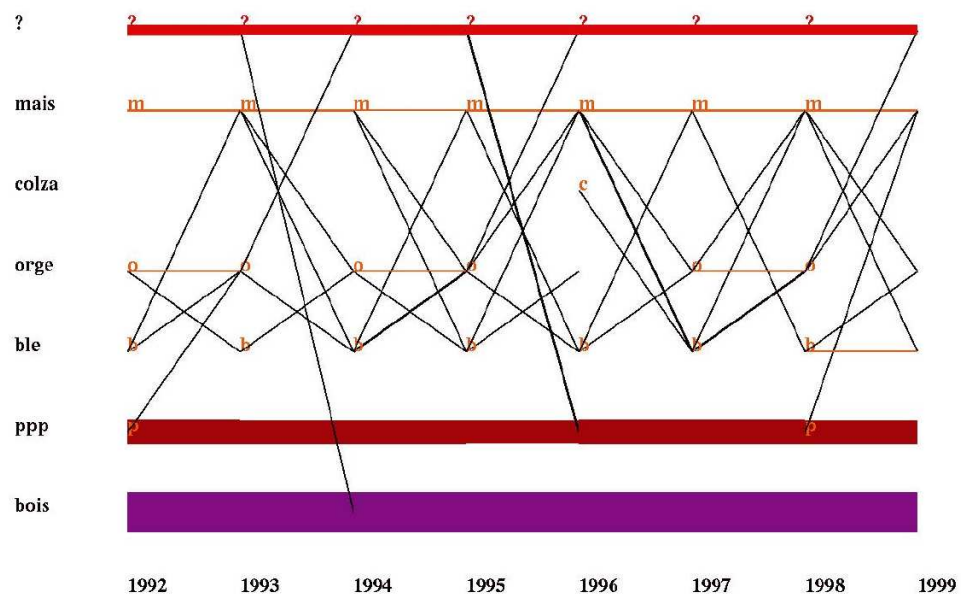


FIG. 9 – Résultats du modèle 2 sur les données des PRA 309 et 310 entre 1992 et 1999.

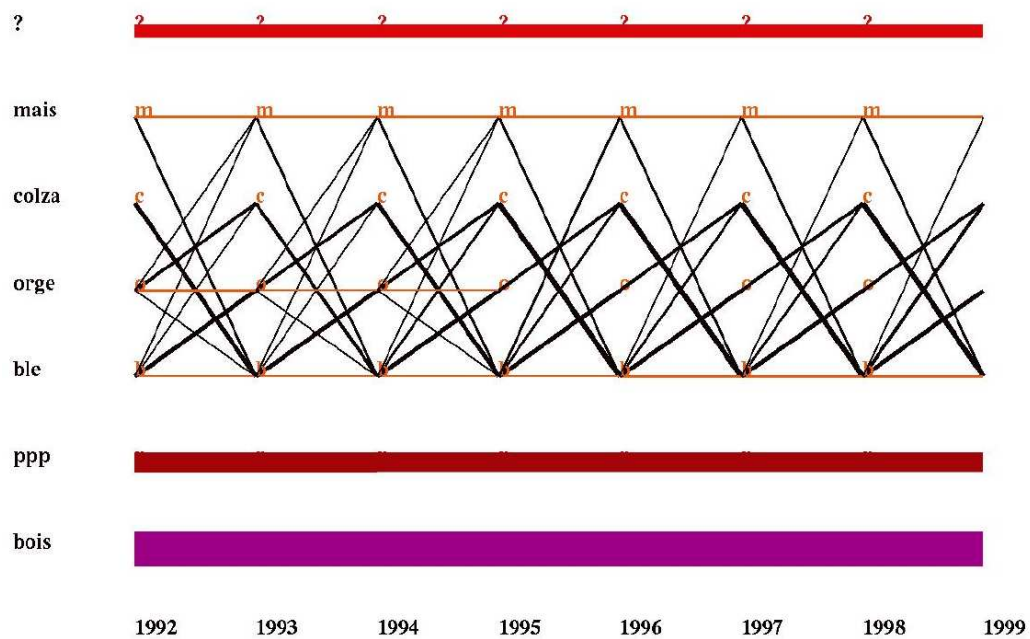


FIG. 10 – Résultats du modèle 2 sur les données de la Lorraine entière entre 1992 et 1999.

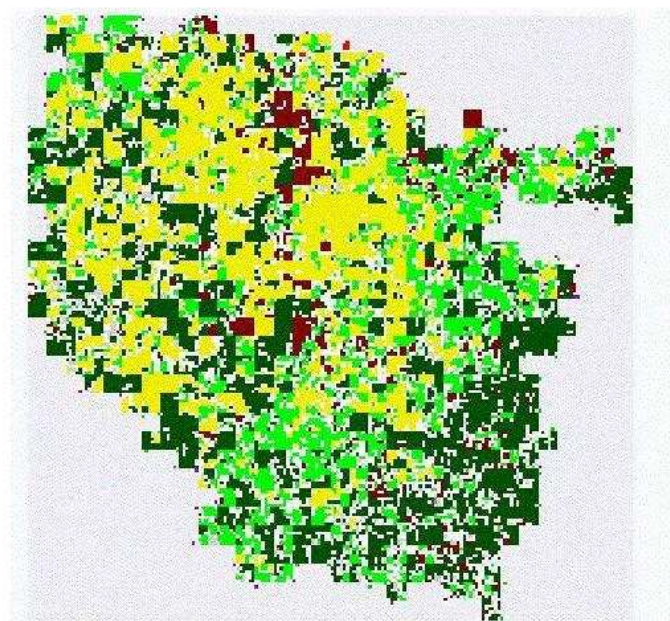


FIG. 11 – *Cartographie des successions culturales en Lorraine. Classification obtenue par un HMM2 à 10 états après 10 itérations.*



FIG. 12 – *Carte de la Lorraine établie dans le cadre du programme Corine Land Cover à partir de données satellitaires (IFEN, 1993). Un point représente 25 ha.*